Hidden Markov Model as a Framework for Situational Awareness

Thyagaraju Damarla

US Army Research Laboratory

Attn: AMSRD-ARL-SE-SA

2800 Powder Mill Road, Adelphi, MD 20783, USA

rdamarla@arl.army.mil

Abstract - In this paper we present a hidden Markov model (HMM) based framework for situational awareness that utilizes multi-sensor multiple modality data. Situational awareness is a process that comes to a conclusion based on the events that take place over a period of time across a wide area. We show that each state in the HMM is an event that leads to a situation and the transition from one state to another is determined based on the probability of detection of certain events using multiple sensors of multiple modalities – thereby using sensor fusion for situational awareness. We show the construction of HMM and apply it to the data collected using a suite of sensors on a Packbot.

Keywords: Keywords: Situational awareness, hidden Markov model, multi-modal sensors, mobile sensor platform

1   Introduction

One of the advantages of using multiple sensors with multiple modalities is to detect various events with high confidence. Situational awareness is achieved based on the sequence of events observed over a period of time. These events may take place in a closed area or on a wide area. In the case of wide area, one would require multiple sensors distributed over the entire region of interest. Situational awareness leads to better response in a timely manner either to mitigate the situation or to take appropriate action proactively rather than reactively. Since the situational awareness is achieved based on the sequence of events observed – hidden Markov model (HMM) [1] is ideally suited. Researchers used HMM for situational awareness for traffic monitoring [2] and learning hand grasping movements for robots [3].

Some of the sensor modalities used in this paper and their possible applications are listed below:

Acoustic sensors: Acoustic sensors are simple microphones used in either an array form or simply used as a single microphone sensor. An acoustic sensor can be used to detect voice or machine made sounds such as the hum of an air conditioning unit or a generator that may be used in a cave environment. The array sensor data can be used to compute the direction of the acoustic signal source.

Seismic sensors: These sensors are similar to the acoustic sensors but sense the vibrations through the ground coupling. These vibrations may be caused by the footfalls of a person walking or a vehicle going on a nearby road. Both the acoustic and seismic sensors do not require line of sight, hence they can monitor the events of interest remotely.

Passive infrared (PIR) sensors: These are simple motion detection sensors. They detect any warm body or object moving in front of it. The amplitude of the sensor depends on the temperature of the object and the distance of the object from the sensor. In some sensors it is possible to determine the direction of the motion.

Chemical Sensor: This is used to detect any hazardous material present in the vicinity. The sensor used has the capability to detect 16 different chemicals.

Magnetic field (B-field) sensors: These magnetic sensors can be used to detect a variety of objects such as magnetic materials carried by people, wires in the walls carrying electric current, generators operating in the vicinity to name a few.

Electric field (E-field) sensors: These are similar to the magnetic sensors. These detect both static and dynamic electric field in the vicinity.

Visible and Infrared (IR) imagers: These are the most robust sensors. The drawback is that they consume more power and require line of sight.

The organization of the paper is as follows: Section 2 presents the problem for situational awareness. Section 3 presents the data analysis of various sensors and events detected by the sensors. Section 4 presents the framework for HMM for the problem. Section 5 concludes the paper.

2   Situational Awareness

Sensor fusion is supposed to lead to a better situational awareness. However fusion of multi-modal data is a difficult thing to do as there are few joint probability density functions for mixed modalities exist. Fusion

		Form Approved OMB No. 0704-0188
	Report Documentation Page	

1. REPORT DATE **JUL 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Hidden Markov Model as a Framework for Situational Awareness	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Army Research Laboratory,Attn: AMSRD-ARL-SE-SA,2800 Powder Mill Road,Adelphi,MD,20783**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
11th International Conference on Information Fusion, June 30 ? July 3, 2008, Cologne, Germany.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**7**	
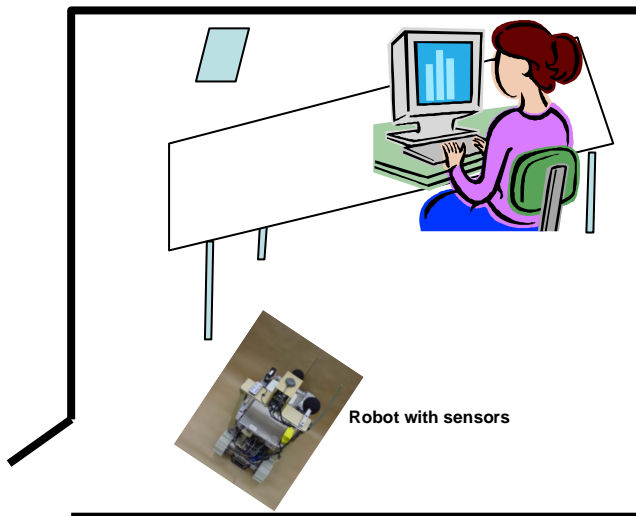
Figure 1: Robot full of sensors monitoring activities in an office room

Figure 2: Robot with different sensors

mostly depends on the application at hand. The problem is further complicated if one has to fuse the events that take place over a period of time and over a wide area for situational awareness. If they are time dependent, relevance of the data observed at different times become an issue. We opted to do fusion of information (decision – probability of detection of an event). In a majority of the cases Bayesian networks [4, 5] are used for fusion. In this paper we use Dempster-Shafer fusion [6, 7] for fusion of multi-modal multi-sensor data. We use the output of fusion to drive the states of HMM.

Some of the situational awareness problems that may be of interest are discussed here. In a situation where we are monitoring a building, we would like to know if there is any activity taking place. In particular, we placed a robot inside an office room (in stealth mode, various sensors will be placed and camouflaged to avoid detection) as shown in figure 1. Figure 2 shows the robot with all the sensors. The goal is to assess the situation based on the observations of various sensor modalities over a period of time in the area covered by the sensor range. We enacted the data collection scenario with several features built-in to observe the happenings inside the office room and assess the situation.

Data Collection Scenario:

- A person walks into the office room – this triggers PIR, B & E-field and seismic sensors.

- She occasionally talks – the acoustic sensor picks up the voice.

- She sits in front of a computer.

- She turns on the computer.

  o B & E-field sensors observe the power surge caused by turning on the computer.
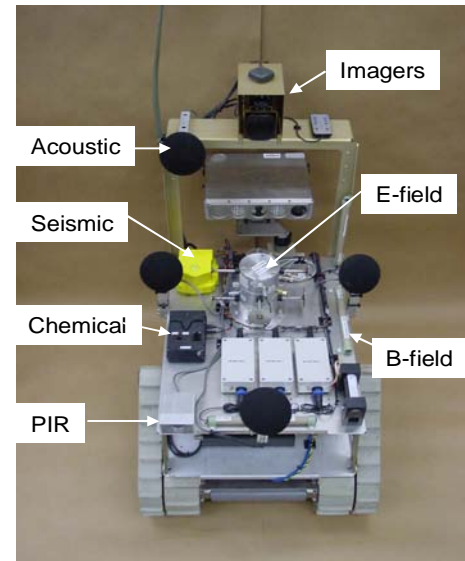
  o Acoustic sensors observe the characteristic chime of Windows turning on.

  o The person's movements are picked up by the PIR sensor.

  o Visible video shows a pattern on the computer screen showing activity on the computer.

  o The IR imager picks up the reflected thermal profile of the person in front of the monitor.

- She types on the keyboard – sound is detected by the acoustic sensor.

- She turns off the computer.

  o Windows turning off sound is observed by the acoustic sensor.

  o The power surge after shutdown is observed by the B-field sensor.

In section 3 we present the data from various sensors and show the events detected by each sensor and also present some of the signal processing done to identify the events.

3 Data Analysis of Various Sensors

We process the data from sensors in order to extract the features corresponding to various events – depending on the situation and application these extracted features will be different even for the same sensor, e.g., voice versus chime.

Acoustic sensor data analysis: In the case of acoustic sensors, we try to look for any human or machine activity – this is done by observing the energy levels in 4 bands, that is, 20 – 250Hz, 251 – 500Hz, 501 – 750Hz and 751 –
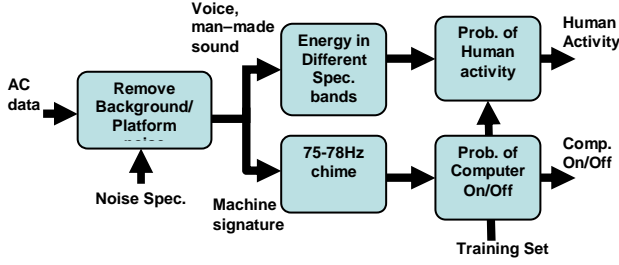
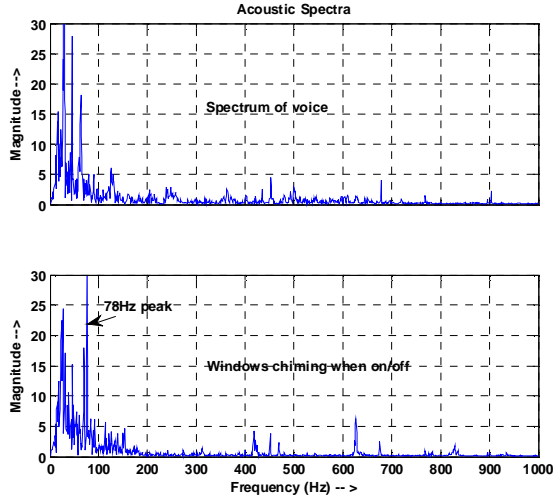Figure 3: Flow chart for acoustic sensor data analysis

Figure 4: Spectrum of voice and computer chime

1000Hz corresponding to voice indicative of human presence. These four energy levels become the feature set and a classifier [8-10] is trained with this feature set collected with a person talking and not talking. The algorithm used to detect a person is presented in [8-10]. For the sake of completeness, the detection algorithm is provided here.

Classifier: Let $\chi = [\chi_1, \chi_2, \cdots, \chi_N]^T$ is a vector of N features, where T denotes the transpose. Assuming they obey the normal distribution, then the multi-variate normal probability distribution of the pattern χ is given by

$$p(\chi) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\chi - M)^T \Sigma^{-1}(\chi - M)\right\},$$

where the mean, M and the covariance matrix Σ are defined as

$$M = E\{X\} = [m_1, m_2, \cdots, m_N]^T$$

$$\Sigma = E\{(X - M)(X - M)^T\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix},$$

and $\sigma_{pq} = E[(x_p - m_p)(x_q - m_q)]$, $p,q = 1,2,\cdots,N$. We assume that for each category i, where $i = \{1, \cdots, R\}$, R denotes the number of classes (in our case $R = 2$, person present and person not present), we know the a priori probability $p(i)$ and the particular N–variate normal probability function $p(\chi|i)$. That is, we know R normal density functions. Let us denote the mean vectors M_i and the covariance matrices Σ_i for $i = 1, \cdots, R$, then we can write

$$p(\chi|i) = \frac{1}{(2\pi)^{N/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{(\chi - M_i)^T \Sigma_i^{-1}(\chi - M_i)}{2}\right\}, \quad (1)$$

where $M_i = (m_{i1}, m_{i2}, \cdots, m_{iN})$. Let us define H_0 and H_1 as the null and human present hypotheses. The likelihood of each hypothesis is defined as the probability of the observation, i.e., feature, conditioned on the hypothesis,

$$l_{H_i}(\chi_s) = p(\chi_s | H_j) \quad (2)$$

for $j = 1,2$ and $s \in \mathcal{S}$, where $\mathcal{S} = \{$acoustic, PIR, seismic$\}$. The conditional probability is modeled as a Gaussian distribution given by (1),

$$p(\chi_s | H_i) = \mathcal{N}(\chi_s; \mu_{s,i}, \sigma_{s,i}^2). \quad (3)$$

Now, (2)-(3) can be used to determine the posterior probability of human presence given a single sensor observation. Namely,

$$p(H_1 | \chi_s) = \frac{l_{H_1}(\chi_s)p(H_1)}{l_{H_0}(\chi_s)p(H_0) + l_{H_1}(\chi_s)p(H_1)}, \quad (4)$$

where $p(H_0)$ and $p(H_1)$ represent the prior probabilities for the absence and presence of a human, respectively. This paper assumes an uninformative prior, i.e., $p(H_0) = p(H_1) = 0.5$. ▼

In the office room scenario, we are looking for any activity on the computer – the Windows operating system produces a distinct sound whenever a computer is turned on or off. This distinct sound has a 75-78Hz tone and the data analysis looks for this tone. The acoustic data process is depicted in the flow chart sown in figure 3 and figure 4 shows the spectrum of the acoustic data when a person is talking and when Windows operating system comes on. The output of the acoustic sensor is P_i, $i = \{1,2,3\}$, corresponding to three situations, namely, (*i*) a person talking, (*ii*) computer chime and (*iii*) no acoustic activity.

Seismic Sensor Data Analysis: We analyze the seismic data for footfalls of a person walking. The gait frequency of normal walk is around 2Hz. We use the envelope of the signal instead of the signal itself to extract the gait
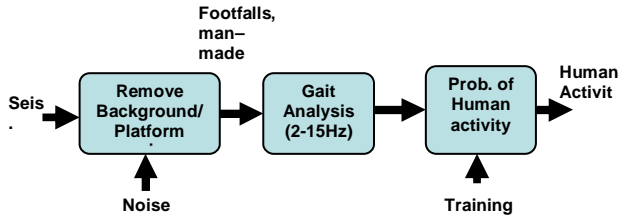
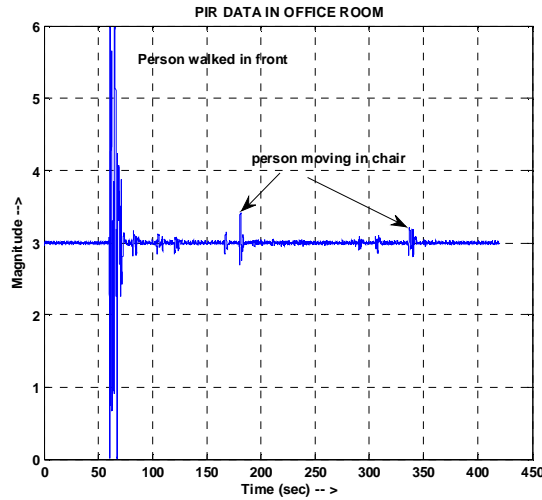Figure 5: Flow chart for seismic sensor data analysis

Figure 6: PIR sensor output

frequency [9, 11]. We also look for the harmonics associated with the gait frequency. Figure 5 shows the flow chart for seismic data analysis. We use the 2-15Hz band to deter-mine the probability of person walking in the vicinity. The seismic sensor provides two probabilities, (*i*) probability of a person walking and (*ii*) probability of nobody present.

PIR sensor data analysis: These are motion detectors, if a person walks in front of them, they will give an output proportional to the temperature of the body and inversely proportional to the distance of the person from the sensor. Figure 6 shows the PIR sensor data collected in the office room. Clearly, one can see a large amplitude when a person walked by the sensor. The smaller amplitudes correspond to the person seated in the chair in front of the computer and moving slightly (note that the chair is obstructing the full view of the person) and only part of the body is seen by the PIR sensor. In order to assess the situation, both seismic and PIR sensor data can be used to determine whether a person entered the office room. The seismic sensor does not require line of sight unlike the PIR sensor – they complement each other.

Magnetic sensor (B-field sensor): We used both Fluxgate and coil magnetometers. The former has low frequency response while the coil magnetometer provides high frequency response. A total of six sensors: three fluxgate magnetometers, one for each direction X, Y, and Z and three coil magnetometers were used. The coil magneto-
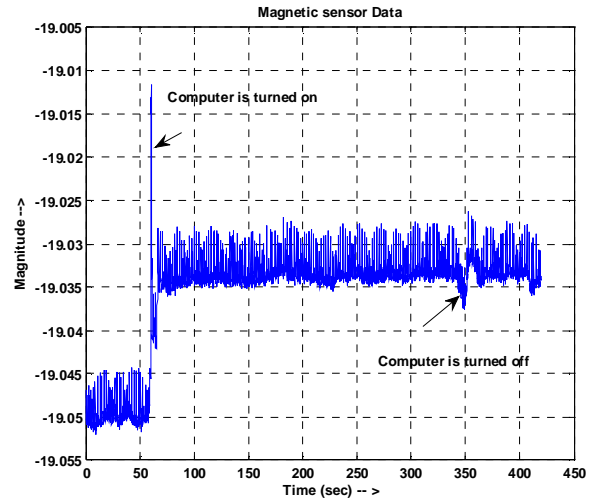

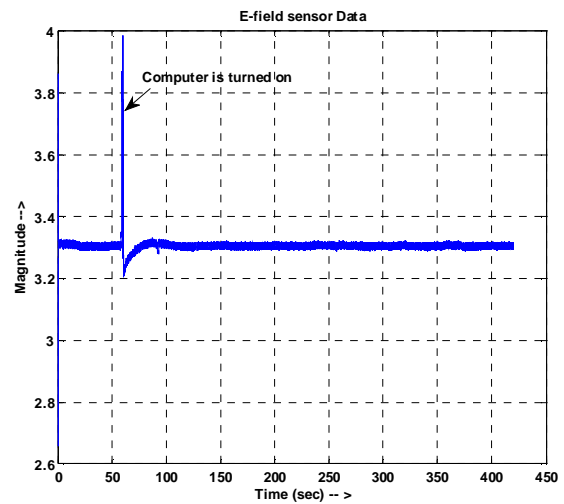Figure 7: Fluxgate magnetometer output in x-axis

Figure 8: E-field sensor output in x-axis

meters are placed in X, Y, and Z axes to measure the magnetic flux in respective direction. Figure 7 shows clearly the change in magnetic flux when a computer is turned on and off. Similar signals are observed in Y and Z axes.

E-Field Sensor data analysis: We used three E-field sensors – one in each axis. The output of X-axis E-field sensor data is shown in figure 8. A spike appears when the computer is turned on in the E-field output, however, we did not observe any spike or change in amplitude when the computer is turned off.

Visible and IR imaging sensors: Several frames of visible and IR images of the office room and its contents are taken over a period of time. In this experiment, the images are used to determine if the computers are on or off and if anybody is sitting in front of the computer to assess the situation. Due to limited field of view of these sensors,

Figure 9: Visible image showing a person in front of computer before it is turned on

Figure 10: IR image frame showing thermal reflection of a person in front of the camera

only a partial view of the room is visible – often it is difficult to observe a person in the room. Figure 9 shows a frame of visible image showing only the shoulder of a person sitting in front of a computer. Figure 10 shows an IR frame showing a thermal image of the person in front of the computer due to reflection. Most of the thermal energy radiated by the person in front of the computer monitor is reflected by the monitor and this reflected thermal energy is detected by the IR imager. The IR imager algorithm processes the silhouette reflected from the monitor – first Hough transform [12] is used to determine the line patterns of an object and then using elliptical and rectangular models to detect a person [13-15] in front of the monitor and provide the probability of a person being present in the room. The visible imager algorithm determines the brightness of the monitor and varying patterns and provides the probability that the computer is on. In the next section we present the framework for HMM.
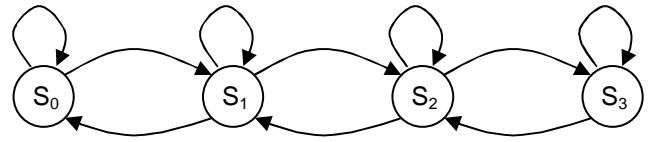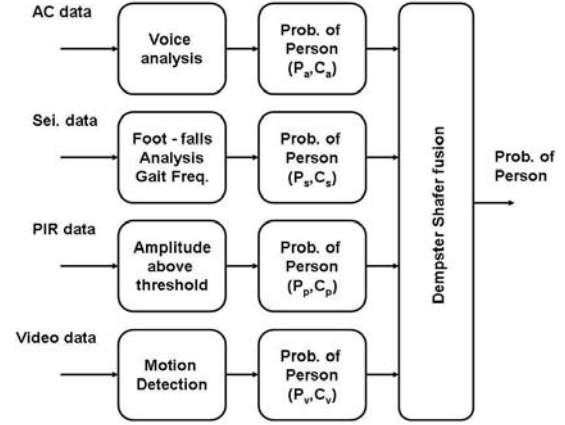

Figure 11: Various states of HMM

Figure 12: Data processing in state S_0

4 Framework for HMM

Based on the situation we are interested in assessing, the HMM is designed with four states as shown in figure 11. The states are as follows:

- S_0 denotes the state when there is no person in the office room,
- S_1 denotes the state when a person is present in the office room,
- S_2 denotes the state when a person is sitting in front of a computer and
- S_3 denotes the state when a computer is in use.

The above mentioned states are just a sample and can be extended to any number based on the situation one is trying to assess on the basis of observations using multi-modal sensors. We now discuss how each state is reached, what sensor data is used and how they are used. This also illustrates that the HMM also achieves the sensor fusion as each state transition is made on the observations of all or a subset of sensors.

State S_0: This is the initial state of the HMM. We use acoustic, seismic, PIR and visible video data to determine the presence of a person. Each sensor gives probability of detection, probability of no detection and confidence level denoted by (P_d, P_{nd}, P_c) as shown in figure 12. These probabilities are fused using the Dempster-Shafer [6, 7] fusion paradigm to determine the overall probability. There will be transition from state S_0 to S_1 if this probability exceeds a predetermined threshold otherwise it

will remain in state S_0. For the sake of completeness, we present the Dempster-Shafer fusion paradigm here.

Dempster-Shafer fusion rule: To combine the results from two sensors (s_1 and s_2), the fusion algorithm uses the Dempster-Shafer Rule of combination [6, 7]: The total probability mass committed to an event Z defined by the combination of evidence represented by $s_1(X)$ and $s_2(Y)$ is given by

$$s_{1,2}(Z) = s_1(Z) \oplus s_2(Z) = K \sum_{X \cap Y = Z} s_1(X) s_2(Y) \quad (5)$$

where $\oplus$ denotes the orthogonal sum and K the normalization factor is:

$$K^{-1} = 1 - \sum_{X \cap Y = \phi} s_1(X) s_2(Y) \quad (6)$$

This is basically the sum of elements from the set of Sensor 1 who intersect with Sensor 2 to make Z, divided by 1 minus the sum of elements from s_1 that have no intersection with s_2.

The rule is used to combine all three probabilities (P_d, P_{nd}, P_c) of sensors s_1 and s_2. The resultant probabilities are combined with the probabilities of the next sensor. ▼

State S_1: This is the state when there is a person in the room. There are three transitions that can take place while in this state, namely, (1) transition to state S_2, (2) transitions back to state S_0 and (3) stays in the same state.

Transition to S_2 happens if any one of the following takes place: (a) if the computer turn on chime is heard, (b) if magnetic and E-field sensors detect flux change and E-field by the respective sensors, (c) if the IR imager detects an image on the monitor and (d) if the visible imager detects changing images that appear during the windows turning on process.

Transition to S_0 takes place if there is no activity on any of the sensors for a duration of 1 minute.

The HMM remain in state S_1 if there is activity in the PIR, acoustic or seismic but not any of the events described for the case of transition to S_2. Figure 13 shows the data processing in each sensor modality.

State S_2: This is the state where a person is in front of the computer. The transition from this state either to S_3 or to S_1 depends on the following: (a) there is keyboard activity or the IR imager detects a hand on the keyboard and the PIR detects slight motion. S_2 to S_1 takes place when the computer is turned off – as detected by acoustic and magnetic sensors.

State S_3: This is the state where the computer is in use. As long as keyboard activity is detected using acoustic and IR imagers the state remains in state S_3, if no keyboard activity is detected for a minute it will transition to S_2.

Data processing in state S_2 is shown in figure 14. Data processing in S_3 is straight forward.
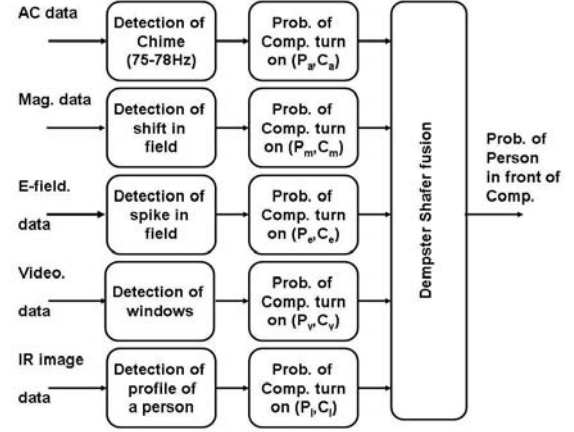

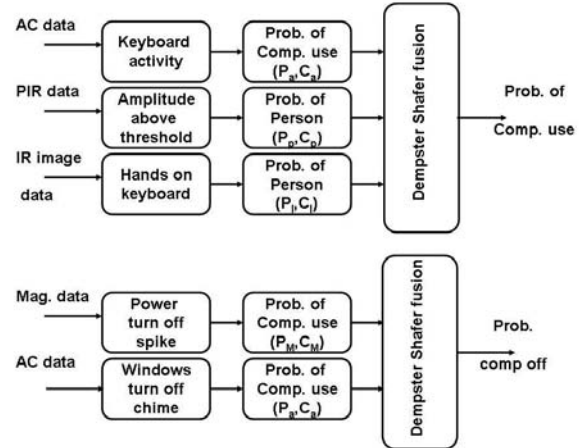Figure 13: Data processing in state S_1

Figure 14: Data processing in state S_2

We discussed what processing is done at each state and how the probabilities are estimated. The transition probabilities of HMM are generated based on several observations with people entering into the computer room, sitting in front of the computer, turning it on, using it for a period of time, turning it off and leaving the office room.

Data processing of various sensors depends on the state of the machine and the confidence levels of various sensor modalities are also changed based on the state of the HMM. For example, in state S_2 the PIR sensor output monitoring a person in a chair produces small amplitude changes as shown in figure 6 – in normal processing those outputs will not result in high probability – however in this case they will be given high probability. In state S_3 the acoustic sensor determines the tapping on the keyboard, this sound is often very light and the sensor is given high confidence levels than normal.

5   Conclusion

We presented HMM as a means to achieve situational awareness using multi-modal multi-sensor data. We also showed that the fusion of information takes place at each

state and that the sensors importance (confidence levels) in each state varies to meet the objectives of situational awareness. The approach can be adopted for any other situation.

References

[1] Lawrence R. Rabiner, "A tutorial on Hidden Markov Models and Selected Application in Speech Recognition", Proc. of the IEEE, VOL. 77, NO. 2, pp. 257-286, February 1989.

[2] Dietmar Bruckner, Brian Sallans, and Gerhard Russ, "Hidden Markov Models for Traffic Observation", Proc. 5th IEEE Intl. Conference on Industrial Informatics, Vienna, Austria, July 23-27, 2007.

[3] Keni Bernardin, Koichi Ogawara, Katsushi Ikeuchi, and Ruediger Dillmann, "A Hidden Markov Model Based Sensor Fusion Approach for Recognizing Continuous Human Grasping Sequences", Proc. 3rd IEEE International Conference on Humanoid Robots, January 2003.

[4] A. Singhal, J. Luo and C. Brown, "A multilevel Bayesian network approach to image sensor fusion", Proc. ISIF, WeB3, pp. 9-16, 2000.

[5] A. Singhal, and C. Brown, "Dynamic Bayes net approach to multimodal sensor fusion", Proc. SPIE Vol. 3209, pp. 2-10, 1997.

[6] D. L. Hall and J. Llinas, *Handbook of Multisensor Data Fusion*, CRC Press: Boca Raton, FL, 2001.

[7] L. A. Klein, Sensor and Data Fusion – A Tool for Information Assessment and Decision Making, SPIE Press, Bellingham, Washington, USA 2004.

[8] T. Raju Damarla, T. Pham and D. Lake, "An algorithm for classifying multiple targets using acoustic signatures", Proc. SPIE Vol. 5429, Orlando, Fl., pp. 421-427, 2004.

[9] T. Damarla and D. Ufford, "Personnel detection using ground sensors", Proc. SPIE Vol. 6562 656205 pp. 1-10, 2007.

[10] Thyagaraju Damarla, Lance Kaplan and Alex Chan, "Human Infrastructure & Human Activity Detection", Proc. of IEEE Fusion Conference, Quebec City, Canada, 2007.

[11] K. M. Houston and D. P. McGaffigan, "Spectrum Analysis Techniques for Personnel Detection Using Seismic Sensors", *Proc. of SPIE,* Vol. 5090, pp. 162-173, 2003.

[12] Hough, P. V. C., "Method and Means for Recognizing Complex Patterns," U.S. Patent 3069654 (Dec. 18, 1962)

[13] Belongie, S., Malik, J., Puzicha, J., "Shape matching and object recognition using shape contexts," IEEE Trans. Pattern Anal. Mach. Intell. 24(4) 2002

[14] Navneet Dalal, Bill Triggs, "Histograms of Oriented Gradients for Human Detection," cvpr, pp. 886-893, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1, 2005

[15] Liming Wang, Jianbo Shi, Gang Song, I-Fan Shen, "Object Detection Combining Recognition and Segmentation," Eighth Asian Conference on Computer Vision (ACCV), 2007.